

Evaluating Oral Competence

Gloria E. ISHIGE

Recently in Japan a teaching syllabus has been required by many schools or opted for by the majority of teachers as it provides a clear idea of the objectives to be achieved for a specific course within a certain time frame. The students also benefit by knowing the overall picture of what they are supposed to study and the goals they are expected to achieve.

The syllabus many of us design when teaching English as a foreign language is a combination of structural, functional-notional and situational features. In oral classes we train our students to be able to manipulate structures, make up their own sentences and use the appropriate language to communicate and interact with other speakers in real-life situations. As we progress in the syllabus, not losing sight of the priorities set in the curriculum and also responding to the students' needs, we are faced with the difficult task of evaluating the students at the end of the term, whether a one-term course or at the end of the first and second semesters of a one school-year course. It is then time to check the students' *achievement* in the course. We also need to know their *proficiency* at that point. A diagnostic overview of the students' achievement and proficiency will give us an insight as to where they are at, where they need to go next and what the teacher should provide them with in the future classes. Whatever the process is, some kind of evaluation has to be done to check the completion of a syllabus in accordance with a program.

When talking about oral ability testing, *achievement tests* immediately come to mind. However, there are also other kinds of tests and objectives in oral language testing. Not to mention very linguistic ones, I would cite the *proficiency tests*. They attempt to measure the foreign-language learner's aptitude in the target language, what the examinee already knows, has mastered and how well he/she can understand and participate in any conversation. The degree of fluency and the accuracy of vocabulary, grammar and pronunciation are also observed. A structured investigation of a learner/speaker's overall ability to handle the language is the aim of this kind of test. In this article I will consider only achievement and proficiency tests as some of the types of oral language testings.

Requirements for Oral Evaluation

In planning an oral test, there are several important points to bear in mind.

Is the test appropriate?

If merely concerned about testing for its own sake, one could pick up a book on

foreign-language testing that offers a multitude of different types of tests and techniques. However, an inappropriate test might result. Allow me to mention a teaching experience I once had.

The teacher of an English conversation course in an intensive English language program suddenly fell ill. Periodic oral tests had been scheduled and the students had been previously notified of them. However, it was on the very day of one of these tests that the teacher became sick. Some faculty members urged that the oral test should be done anyway, since the school was well known for its reputation and long track record in following through on its teaching programs. But all of the English conversation teachers who had or were teaching the same or a similar course were engaged, so I was requested to take over in the emergency. I was unaware of the course syllabus and the only information provided was the academic year the students were in. I agreed to substitute for the sick teacher for humanistic reasons, but also thought that the improvised test would provide me with interesting teaching insights. Different aspects of speaking ability were measured: fluency, accuracy (vocabulary, grammar, intelligibility), pronunciation and intonation. As might be expected, the grades of this unplanned test turned out to be much lower than the average of previous and subsequent periodic tests. Note Table 1 below.

Table 1. Comparison of average grades on test by substitute teacher and periodic tests

Speaking Skill	Test by Substitute Teacher (average grade)	Periodic Tests (average grade)
Comprehension (necessary to measure oral competence)	48	72
Fluency	34	68
Accuracy - Vocabulary	32	70
- Grammar	45	62
- Intelligibility	52	75
Pronunciation, intonation	47	65

I observed and reported to the school administration a great variation in individual performance on the test. The score gap between the good students, who composed a small group and the poor ones who represented the majority was very large. The heterogeneous oral ability of the students, as later brought home to me, was confirmed in the test results. Only one-fourth of the speaking goals targeted in the test seem to have been realized. An unplanned, inappropriate test for that particular class resulted in very low marks, despite the fact that they might have scored very highly. I was informed that this program's periodic tests did not aim to simply measure personal oral

competence, but the achievement of each student, especially the poor ones after certain conversation patterns or communicative grammar points were practiced. Upgrading the oral ability of poor speakers was the ultimate goal. The oral test I gave was not appropriate to that program as it ended up gauging only each individual's general speaking skill.

Oral tests, like any other, can be considered appropriate if planned and prepared in advance taking into account the syllabus, curriculum and general objectives of that specific course.

Is the test valid?

When a teacher-tester prepares or selects an already made test, it should look like a good one as a whole, and satisfy both the examiner and the student being tested. The content should reflect the syllabus it was based on. Its format and procedure should allow a duplicate performance in subsequent real-life situations — the oral production of the assigned or requested task should sound natural and be practical. Lastly, regular assessments of tests have proved that a valid test shows similar results to previous ones that have also been validated.

A testing situation is, admittedly, always somewhat artificial. "Oral interviews, discussions or role-plays between students or with the teacher are not direct measures," is the criticism we hear from some linguists. Would talking with nonteachers or students from other schools be a more "authentic" situation in which we could make direct measures? Another criticism has been made of the testing approach in which students are divided into small groups and asked to discuss among themselves and later report on solutions to a problem, or summarize a play or short movie they have seen in class. The claim is that students "Japanize" their speech when communicating with their usual partners. Sharing a common language and being familiar with each other encourages them to communicate in Japanese. I have conducted this type of test myself many times as well as undertaken this activity in class. As I monitor the students when they are discussing or preparing for the presentation, I observed that they feel mutually challenged and stimulated to complete the task in the best way. Though their communication may not be in very accurate and fluent English, at least English was used.

Is the test doable?

Oral testing may mistakenly lead us to think of speed, cost and physical energy economy since assessment is carried out mostly verbally. However, we cannot take these premises for granted, nor can we completely subject ourselves to them.

Considering the variety of oral tests, they must be cost-effective as far as the use of material resources is concerned. And the use of human resources has to be even more carefully monitored. Although more than one examiner situation is the ideal for subjective tests such as oral interviews, it is often impossible in regular academic course pro-

grams. Manpower requirements and feasibility are often crucial points. In academic oral courses, the students do almost all the talking, so more than one testee could be examined in the same amount of time by a single examiner, depending on the nature of the test. The time consumed would average fifteen minutes at the longest. However, in case of oral language proficiency tests given to the general public, it is advisable that the examiners who are teaching similar levels and using similar approaches or materials work together when elaborating the test as the number of examinees is usually much larger. An agreement on the rating scale will enable them to be tested together in a shorter time frame. Examiners would then be able to endure longer periods of testing and experience less fatigue, as studies have shown. The teacher should carefully figure out the needed time to administer the test and plan the test schedule beforehand. Scoring might therefore not require so much time.

The students' psychological reaction and behavior should not be ignored. Will the test content seem unfair to the students? Will he/she feel comfortable enough to undertake the required task? Will the student be able to adopt the appropriate emotional attitude to elaborate and perform the task? We should also consider the factor of physical fatigue. The length of the test, the testing environment (space, light, room temperature, comfort (desk, chair) may affect the student in different ways and lead to variable results.

The practicability of an oral test is also measured by its comparability with previous test scores. Tests should share some common features to allow later comparisons which will enable a more accurate evaluation of the students and gauge the components of the course as well.

Is the test reliable?

Reliability is probably the most controversial and complex issue in oral language testing. We have to consider reliability in terms of the test itself and the scoring.

In the planning stage and after elaborating an oral test, one should make sure the questions or formulated tasks are comparable in difficulty and require all the examinees to perform basically the same or similar types of things. Not only the questions but also the instructions should be simple and clear enough to avoid misinterpretation. At times, unfortunately, instructions present more difficulties than the test items themselves. I have personally come across some very confusing ones myself. In most cases they were too vague.

When a test is composed beforehand, the tester objectively formulates it expecting some kind of task completion. If, on the other hand, the teacher subjectively decides at each time, in an improvised way, what to ask and whether to continue or not because the student has shown a good command of the language or solved the tasks satisfactorily, the student's test may unfairly end too soon. The teacher may be erroneously assuming the examinee can handle easier tasks as more difficult ones were done well. The result may be unreasonable overgrading or undergrading even though the teacher

may claim impartiality.

A carefully structured test as well as the use of a checklist and a rating scale may again be very helpful to achieve test objectivity and reliability when evaluating the students. Subjectivity may negatively influence a teacher if there are not preset criteria. For example, an examiner may become a victim of the so-called "halo effect," giving grades higher or lower than the student deserves. Also, less experienced and or not professional-minded instructors may make their personal relations with students too much the key, becoming, at times, extremely subjective and emotional when scoring their performance.

Where oral proficiency tests given to unknown testees, the fact that the testers are not emotionally involved with the examinees before the tests, more objectivity and more appropriate evaluation is assumed. On the other hand, does less interaction with testees before or during the test really make the test more reliable? Teacher-examiners, including myself, note that sometimes "good students" in the class perform less satisfactorily on tests, whereas "quiet students" at times surprise us by doing much better than expected. Pairing a "quiet student" with a shyer partner in a test may encourage the "quiet" one to be more motivated to perform. I have observed on one occasion that the "quiet" examinee was sensitive to that situation, and very naturally tried to make it easier for the "shy" partner by initiating things and taking the lead in the assigned task. This "behavioral improvement" resulted in a better linguistic performance. Individual differences, personality and its variables should always be taken into account when attempting to obtain accurate results.

Constant and careful observation of student performance in class, interaction and involvement of the teacher in class activities do help provide a fair knowledge of student ability in oral work outside a test situation as well. Consequently, a much more reliable analysis and evaluation are possible.

Types of Testing and Procedures

Given the various problems and conditions one must consider when administering a test, teacher-examiners are left with a difficult task. Objectivity seems to be the first requisite of a good test. An objective test allows the same scoring by several and different examiners where fair grades can be expected. *Multiple-choice tests* have been frequently used to measure speaking ability as an objective approach. Through the *directed speaking test* method a high degree of objectivity is within reach since the response possibilities are narrowly controlled. Model answers are provided in the test instruction phase. All other responses are considered incorrect. This method has been used especially with beginners and non-fluent speakers the same way they have practiced language structures in classroom drills. For example, the examinee is requested to convert sentences from affirmative to negative or interrogative forms, ask direct questions when an indirect statement is given, e.g., "Ask me how long it took me to bake this cheese pie."

Time measure procedure can be used for grading oral performance. This procedure can be applied to measure the amount of time required to formulate questions on a passage, summarize a passage, elaborate or add comments to a passage, or perform a rather complex oral task. It also serves to gauge the number of short, simple messages that a student can compose orally. Time measure can again be used when conducting tests to assess oral accuracy in terms of grammar, pronunciation, etc.

Individual interview tests conducted by the teacher-examiner is the most common type of oral testing. The examiner can measure fluency in tasks like the following:

1. Examiner reads a brief statement and asks the student to formulate questions about it. For example: "My family decided to host a student from Switzerland."
2. Examinee is asked to report, for example, on a field trip, club activities as a student, a favorite movie, a recent visit to an amusement park.
3. Examinee is asked to state his/her opinion on a current problem.
4. Examinee is requested to comment on something the examiner is going to say.
5. A personal problem is stated and the examinee is asked to provide advice and relevant solutions.
6. Examiner shows a picture, a series of pictures or situational pictures and asks the testee to describe, comment or make up a story with the pictures in mind.

To test knowledge of appropriate functions and notions, I suggest tasks such as:

1. Ask the student what he/she would say in a situation presented. For example:
 - a. It's your friend's birthday. You have bought him/her a present.
You are standing in front of your friend holding the gift.
 - b. You are at a party. You are talking to someone you have just met.
You are a little bored. You want to escape.
2. Make a statement and ask the student to give or make an appropriate statement.
For example:
 - a. My sister is very sick in the hospital.
 - b. Little Bob broke Grandma's favorite vase. This is the third time he broke something precious to her.
 - c. Do you mind if I open the window?
3. Ask the student to change from informal to formal language. For example:
(Sample Reply)
 - a. How's it going? (a. How are you?)
 - b. I've gotta get going. (b. I'm sorry, I must go now.)

To test pronunciation, intonation, or the general speed of speech through reading, ask the examinee to:

1. Read a familiar dialog, passage or poem.
2. Read an unfamiliar dialog or a new material.

Individual interview tests conducted by the examinees are motivating since they can lead the role-play by pretending to be interviewers. But this sort of testing is often considered unfeasible because they require the presence of a stranger who is to be inter-

viewed. When oral proficiency tests given to unknown examinees are carried out by a group of examiners, oral tests of this type are more feasible since one examiner can act as the interviewed person. Giving oral tests during an academic year may present a difficult but not impossible situation. If the whole class as a group were to function as interviewers, one might fear that only the more fluent students will do all the talking. However, in my experience, advising the students that this is a test and they need to show their speaking ability will make this technique successful and obtain interesting results. Written materials (like forms to be completed by each interviewer) have to be prepared. The requested information will direct the student-interviewer to formulate questions. To help the observing teacher-examiner in the evaluation, it could be stipulated that when a student's question is structurally incorrect, the interviewer need not give a reply.

One variation of individual interview tests conducted by the examinee might be to have the student interview the teacher-examiner. In this test scenario, at least five different written questionnaires could be prepared containing items on which the examinee can formulate questions and interview the examiner. Content or vocabulary familiar to the student should be incorporated in the tasks to allow a fair assessment. Four copies of each questionnaire should be then made, giving the impression that there are twenty different forms. The twenty forms should be placed face down on a table. The examinees come into the classroom or the teacher-examiner's office individually for the oral test. The numerous forms lying on the table gives the examinee the impression that there are many different questionnaires, discouraging testees from giving fellow testees details on the test tasks. The examinee picks one form and then interviews the examiner. To make the testee more comfortable, tape recorders can be used to record the interview. (Constant note-taking of the examiner may unnerve or distract the student.) There is no doubt this procedure provides objective assessment of a person's oral competence because the examiner can review the interview tapes repeatedly. The disadvantages are the time and effort required to go over all the recordings.

Pair-communication tests are when two examinees are asked to perform a task by communicating with each other. A variety of tests can be devised. An interview setting like the one described above can be done by two examinees. One of the testees plays the interviewed person. To provide equal opportunity for both students to interview and be interviewed, another variation can be devised. Two different questionnaire forms are given to the testees or drawn by them. After completing the interviews, the interviewers are asked to report the information collected. This kind of task demands the use of different basic structures, vocabulary and expressions, which are either familiar or recently learned in class.

Discussion topics covered in class or current topics, in case of proficiency tests, can also function as stimuli. Strips of paper are prepared with discussion topics. Each pair draws one and discussion is carried out.

The ability to handle situations in conversational English can be tested preparing a

number of stimuli with instructions like, for example:

“A has finished eating in a restaurant. A wants to pay for the meal.

A calls B, the waiter/waitress. A pays B directly.”

Or something like:

“A has been invited to B’s house for dinner. B is A’s superior. B welcomes

A at the door. B offers A a seat and something to drink before dinner.”

Depending on the examinees’ ability level and the examiner’s criteria, time can be granted to prepare and practice the dialog. Student initiative and a certain amount of ingenuity can be expected in role-plays. Situational conversations test the ability to use proper formal or informal communicative expressions, hesitation words, facilitators, as well as the ability to add comments, change the direction of a conversation and other linguistic features, as the examinees engage in role-playing.

Small-group-communicative tests of three or four testees can also be designed for topic discussions, discussion followed by reports or role-plays. The disadvantage compared with pair-communication tests lies in the shortened time each speaker is going to perform his/her part and the limited oral data to assess the performance in an objective and fair fashion.

Evaluating Oral Competence

Any language course is designed with a syllabus of subskills and linguistic features to be practiced and mastered. Actually, any time a student attempts a given task, he/she is taking a test. As the teacher assigns homework and oral presentation on a weekly basis and gives periodic oral review exercises, keeping records of students’ class performance will help the instructor judge more fairly each student’s actual oral ability because the student-testee may sometimes become extremely nervous in a test situation.

In academic programs, an evaluation sheet with itemized objectives help the instructor analyze and measure the students’ progress and achievement. This form can be completed every class, once every two weeks, after each task assigned for review and mastery is performed or only after an objective is accomplished. Grading, a simple check-mark or date can be entered. An evaluation sheet similar to the one below can be redesigned to suit the teacher’s planning, criteria, and available time.

I have tried to show that it is possible to rate the achievement of language skills or end goals. Since language is the result of an interaction of elements, it would be incorrect to say that we can determine oral competence by how much language he/she knows. The ability to naturally understand, react to cultural-linguistic insights and communicate in acceptable and intelligible language will determine the speaker’s oral competence. It can be, therefore, viewed as the result of an accumulation of linguistic achievements. When evaluating oral performance, we are at times left puzzled, wondering whether the examinee is truly competent or not. Each student has a certain personal appeal and linguistic ability that might compromise our estimation of his or her capabilities. One speaker may appear to be much better than the other. To verify consistent ability, I

see the need for both kinds of achievement and proficiency testing.

Objective	Student: Aoki, Mia	Student: Ito, Keiko	Student:
1. Interviews			
a. Fluency			
b. Accuracy			
2. Oral Reports			
a. Fluency			
b. Accuracy			
3. Role-plays			
a. Fluency			
b. Accuracy			
4. ...			
5. ...			

Conclusion

Communicative oral testing clearly reflects the complexity of the tested language itself. Any two students, two examinees, or two testing situations cannot be the same. Each individual responds differently and the mistakes made cannot be rigidly categorized. However, if effective procedures are devised and good testing planned and prepared, oral testing can be successful and done satisfactorily in a less time consuming way.

In academic courses, oral testing cannot be a one-time evaluation. Many failures on the part of a student-testee may be due to personality factors, sickness, problems at home and other circumstances that affect any human being. Therefore, as oral evaluation cannot be done mechanically because of its highly subjective nature, only extensive all-round oral work in class with careful observation of a student's class-to-class performance will enable fair evaluation.

There is no doubt teaching and learning are more important than testing in oral programs. Nonetheless I believe oral tests are absolutely necessary. I have observed in the oral courses I have conducted and I can testify that when perseverance and diligence informed a student's class work, the success he/she experienced on tests of performance was the most motivating component in his/her learning process. On such occasions, an encouraging, congratulatory comment, or a smile from the teacher enhances a learner's self-image and builds confidence for future efforts to learn the language.

Bibliography

1. Brumfit, C.J. and K. Johnson. 1979. *The Communicative Approach to Language Testing*. Oxford: Oxford University Press.
2. Rea-Dickins, P. and K. Germaine. 1992. *Evaluation*. Oxford: Oxford University Press.
3. Savignon, S.J. 1983. *Communicative Competence: Theory and Classroom Practice*. Addison-Wesley Publishing Company.
4. Spolsky, B. (3rd ed.) 1992. *Conditions for Second Language Learning*. Oxford: Oxford University Press.
5. Vallette, R.M. (2nd ed.) 1977. *Modern Language Testing: A Handbook*. New York: Harcourt Brace Jovanovich.